```
1  import pandas as pd
2  df = pd.read_csv('/content/Retail_Transaction_Dataset.csv')
3  df.head(10)
```

	CustomerID	ProductID	Quantity	Price	TransactionDate	PaymentMethod	StoreLocation	ProductCategory	DiscountApplied
0	109318	C	7	80.079844	12/26/2023 12:32	Cash	176 Andrew Cliffs\nBaileyfort, HI 93354	Books	18.677
1	993229	C	4	75.195229	8/5/2023 0:00	Cash	11635 William Well Suite 809\nEast Kara, MT 19483	Home Decor	14.121
2	579675	A	8	31.528816	3/11/2024 18:51	Cash	910 Mendez Ville Suite 909\nPort Lauraland, MO...	Books	15.943
3	799826	D	5	98.880218	10/27/2023 22:00	PayPal	87522 Sharon Corners Suite 500\nLake Tammy, MO...	Books	6.686
4	121413	A	7	93.188512	12/22/2023 11:38	Cash	0070 Michelle Island Suite 143\nHoland, VA 80142	Electronics	4.030
5	463050	D	3	54.093152	8/15/2023 4:24	Cash	8492 Jonathan Drive\nNorth Robertshire, TN 67532	Electronics	10.888
6	888163	D	7	13.121937	12/26/2023 5:32	PayPal	USNV Harrell\nFPO AA 62814	Clothing	16.295
							489 Juan Loop Apt		

Next steps: [Generate code with `df`]   [View recommended plots]

```
1  df.shape
```

```
(100000, 10)
```

```
1  # Summary statistics for Price, DiscountApplied(%) and TotalAmount
2  df[['Price', 'DiscountApplied(%)', 'TotalAmount']].describe()
```

|	Price	DiscountApplied(%)	TotalAmount
count	100000.000000	100000.000000	100000.000000
mean	55.067344	10.020155	248.334955
std	25.971567	5.779534	184.554792
min	10.000430	0.000046	8.274825
25%	32.549474	5.001013	95.163418
50%	55.116789	10.030353	200.368393
75%	77.456763	15.018367	362.009980
max	99.999284	19.999585	896.141242

```
1  # Correlation between Price, DiscountApplied(%) and TotalAmount
2  df[['Price', 'DiscountApplied(%)', 'TotalAmount']].corr()
```

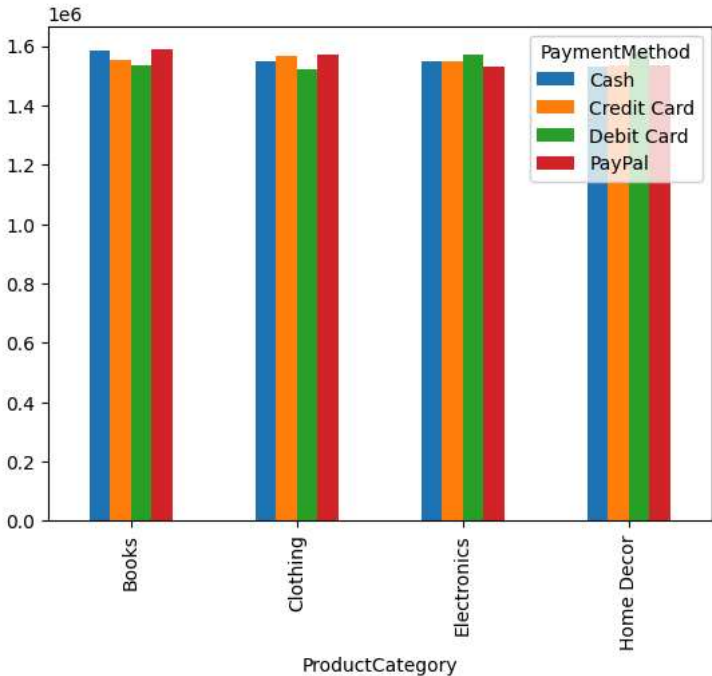|	Price	DiscountApplied(%)	TotalAmount
Price	1.000000	-0.000068	0.636294
DiscountApplied(%)	-0.000068	1.000000	-0.087858
TotalAmount	0.636294	-0.087858	1.000000

```
1  #Total Amount by Payment Method
2  df.groupby('PaymentMethod')['TotalAmount'].sum().plot(kind='pie', autopct='%1.1f%%')
```

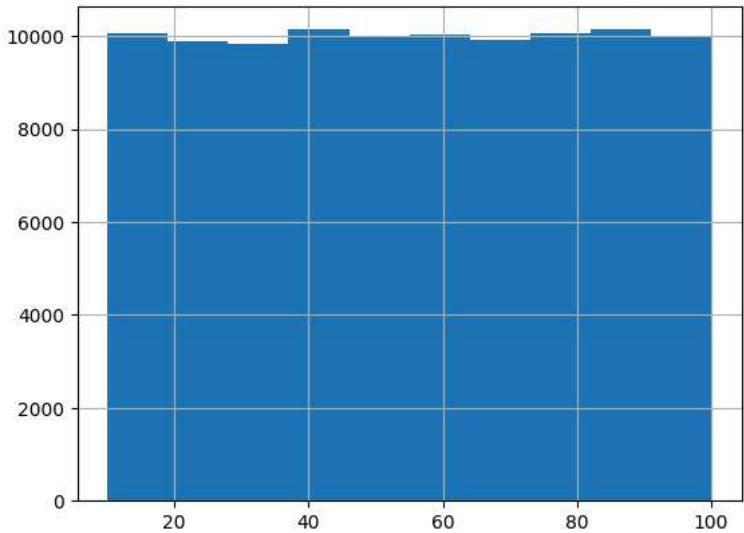<Axes: ylabel='TotalAmount'>
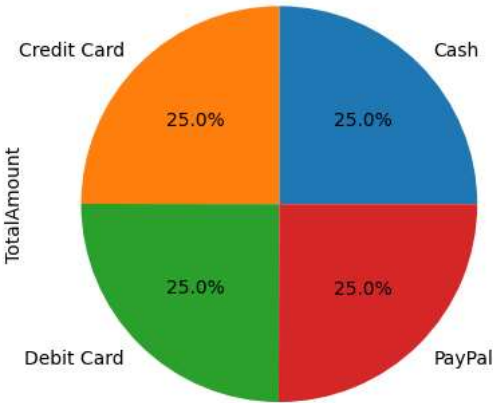

```
1    #Total Amount Spent by Product Category and Payment Method
2    df.groupby(['ProductCategory', 'PaymentMethod'])['TotalAmount'].sum().unstack().plot(kind='bar')
```

<Axes: xlabel='ProductCategory'>
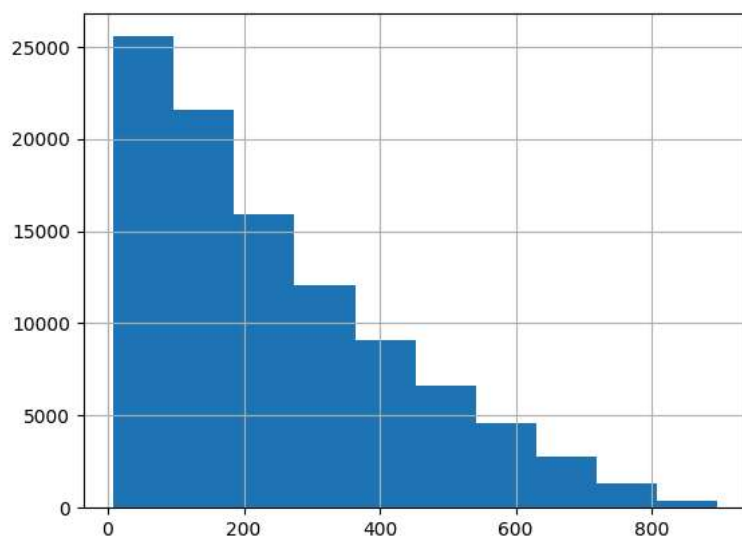


```
1 # Distribution of Price
2 df['Price'].hist(bins=10)
```
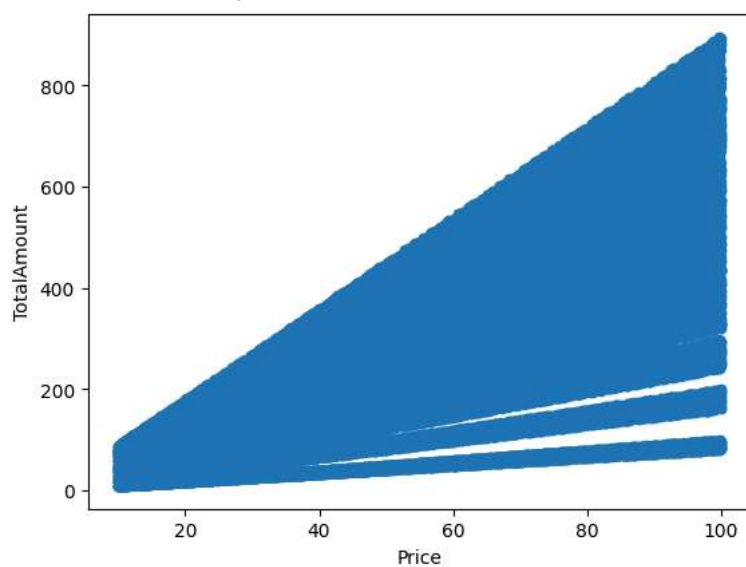
<Axes: >

```
1 # Distribution of TotalAmount
2 df['TotalAmount'].hist(bins=10)
```

<Axes: >


```
1 # Scatter plot of Price vs. TotalAmount
2 df.plot.scatter(x='Price', y='TotalAmount')
```

<Axes: xlabel='Price', ylabel='TotalAmount'>


```
1  # Scatter plot of DiscountApplied(%) vs. TotalAmount
2  df.plot.scatter(x='DiscountApplied(%)', y='TotalAmount')
```

```
<Axes: xlabel='DiscountApplied(%)', ylabel='TotalAmount'>
```



```
<Axes: xlabel='DiscountApplied(%)', ylabel='TotalAmount'>
```